Bak–Sneppen type models and rank-driven processes

Michael Grinfeld,* Philip A. Knight,† and Andrew R. Wade‡
Department of Mathematics and Statistics
University of Strathclyde,
26 Richmond Street, Glasgow G1 1XH, UK
(Dated: October 20, 2011)

The Bak–Sneppen model is a simple stochastic model of evolution that exhibits self-organized criticality and for which few analytical results have been established. In the original Bak–Sneppen model and many subsequent variants, interactions among the evolving species are tied to a specified topology. We report a surprising connection between Bak–Sneppen type models and more tractable Markov processes that evolve without any reference to an underlying topology. Specifically, we show that in the case of a large number of species, the long time behaviour of the fitness profile in the anisotropic Bak–Sneppen model can be replicated by a model with a purely rank-based update rule whose asymptotics can be studied rigorously.

PACS numbers: 02.50.Ga; 05.40.Fb; 05.65.+b; 87.23.Kg
Keywords: Bak–Sneppen model; thresholds; rank-driven Markov processes; self-organized criticality; evolution

In [1], Bak and Sneppen introduced a very fruitful and simple model of evolution that exhibits interesting dynamics but has proved surprisingly hard to analyse. The classical Bak–Sneppen (BS) model is a simple stochastic coarse-grained model of evolution of an ecosystem consisting of a fixed number N of evolutionary niches organised in a ring. Each niche is occupied by a species with a particular *fitness* value in $[0, 1]$. Direct inter-species interactions (predation, competition, etc.) occur only between species in neighbouring niches. The dynamics of the system is driven by the removal (extinction) of the least fit species in the entire system, whose niche is taken over by a new species; the extinction of the least fit species induces changes in the fitnesses of the species in the two neighbouring niches. In this letter, we show that a process whose update rule is defined solely in terms of the *ranks* of the fitness values, without any reference to topology of interactions, exhibits asymptotic behaviour similar to the (anisotropic version of the) BS model as well as self-organized criticality [2]. We call processes of this type *rank-driven processes* (RDPs) and analyse them rigorously in [3]. RDPs are of independent mathematical interest and can be used to define new evolution models. Despite the considerable impact of the Bak–Sneppen model on the physics community and beyond [4], so far only a small number of rigorous results on the BS model have been obtained, such as those of Meester and Znamenski [5] on the non-triviality of the steady-state distribution. In this letter, by exploiting the tools for analysis of RDP models developed in [3], we provide an approach for establishing new results in this active area.

The BS model [1] is a discrete-time Markov process which advances every time there is a species extinction event. Each species occupying the N niches is initially assigned a fitness $x_k \in [0, 1]$ ($k \in \{1, \ldots, N\}$) chosen independently from the uniform distribution on the unit interval, $U[0, 1]$. At each step of the algorithm, we choose the smallest of all the x_k, x_{kmin} say, and replace x_{kmin} and its two nearest neighbours $x_{kmin\pm1}$ (indices calculated modulo N) by new independent $U[0, 1]$ random numbers. In simulations with large N, the marginal distribution of the fitness at any particular niche is seen to evolve to a $U[s^*, 1]$ distribution, with $s^* \approx 0.667$.

A number of variants of Bak and Sneppen's original model have been introduced which evolve according to different topological criteria. One simple variant is the *anisotropic* Bak–Sneppen (aBS) model, in which, in addition to the least fit species, only its *right-hand* nearest neighbour is replaced. The aBS model is the main focus of this letter because while it simplifies calculations, it preserves the key qualitative phenomena of the original BS model. For example, the aBS model also gives rise (according to large-N simulations) to a threshold value $s^* \approx 0.724$ [6]. One contribution of the present letter is to propose a characterization for s^* in terms of ostensibly simpler quantities associated with aBS. Our arguments connecting rank-driven processes to aBS apply to BS too.

Another variant on the BS model which eliminates topology is the mean-field version analysed in [7–9], in which one replaces the smallest fitness and $K - 1$ randomly chosen other ones. Below we show that such models fall within the RDP framework.

Consider a process in which at each update the species with the smallest fitness and the R-th ranked fitness are replaced, where R is a random variable on $\{2, 3, \ldots, N\}$ ($R = N$ corresponding to the largest fitness) sampled independently from a distribution $P[R = k] = f_N(k)$

where $f_N(k) \geq 0$ and $\sum_{k=2}^{N} f_N(k) = 1$. This is an example of a rank-driven process. The complexity of this RDP is intermediate between that of the aBS model and the mean-field model of [7–9]; the latter is the special case of a RDP with $f_N(k) = \frac{1}{N-1}$ for all $k \in \{2,\ldots,N\}$. The RDP has the advantage over aBS that it can be analysed rigorously. We have strong numerical evidence that for a judicious choice of $f_N(k)$ this simpler model can replicate the asymptotic behaviour of aBS.

Specifically, one can choose $f_N(k)$ to be $f_N^{\mathrm{aBS}}(k)$, the *empirical* distribution of the rank of the second chosen site in aBS: if we let $P(k,M)$ be the number of times the k-th ranked element, $k \geq 2$, is the right neighbour of the smallest element in M iterations of the aBS algorithm,

$$f_N^{\mathrm{aBS}}(k) = \lim_{M\to\infty} \frac{1}{M} P(k,M). \tag{1}$$

Heuristically, we expect that given suitable ergodicity properties for the aBS Markov process on the uncountable state space $[0,1]^N$, this limit will exist with probability one.

Little is known analytically about f_N^{aBS}, due to the difficulty of the aBS model, but f_N^{aBS} can be accurately numerically computed. Figure 1 shows simulation estimates of $f_N^{\mathrm{aBS}}(k)$ for small values of k and different values of N.

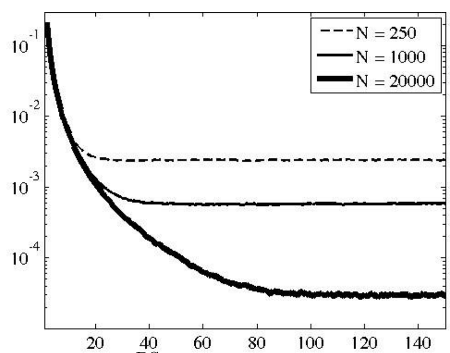

FIG. 1. Plot of $f_N^{\mathrm{aBS}}(k)$, $k \in \{2,\ldots,150\}$ for $N = 250, 1000, 20000$.

The RDP with $f_N = f_N^{\mathrm{aBS}}$ computed numerically replicates many aspects of the dynamics of aBS. Specifically, rigorous results show that such an RDP exhibits a (large-N, long-time) threshold at some s^* which is an explicit function of f_N: for the particular choice $f_N = f_N^{\mathrm{aBS}}$, using our numerical estimates for f_N^{aBS} gives rise to a value for s^* which is very close to the value computed directly from simulations of aBS.

In [3], by considering the random walk associated with an RDP defined in terms of f_N, we show that a crucial quantity is

$$\alpha = \lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=2}^{n} f_N(k), \tag{2}$$

assuming that the N-limit exists. Here $\alpha \in [0,1]$ measures the "atomicity" of f_N as $N \to \infty$. For example, for the mean field aBS of [7–9] one has $\alpha = 0$, while if we always replace the smallest and the second smallest elements, $\alpha = 1$. The main result of [3] is that the threshold in the limiting ($N \to \infty$) stationary distribution of x values in the RDP is given by

$$s^* = \frac{1+\alpha}{2}. \tag{3}$$

A second result of [3] shows that the limiting marginal distribution at stationarity is $U[s^*,1]$, where s^* is given by (3), provided that the selection distribution f_N is "eventually uniform" in the sense that

$$f_N(k) \approx \frac{1-\alpha}{N} \tag{4}$$

for k sufficiently large. This condition is satisfied with $\alpha = 0$ for the mean-field aBS model, showing that the limiting distribution is indeed $U[1/2,1]$ in that case, as indicated by [8].

From the results of Figure 1 we see that for a given N, $f_N^{\mathrm{aBS}}(k)$ decays rapidly for small k before settling down to a uniform value. In fact, it appears that there is a constant C such that $f_N^{\mathrm{aBS}}(k) = C/N$ for large enough k. Thus the numerical evidence supports the eventual uniformity condition (4). Hence $\alpha = 1 - C$. Numerical results give $\alpha \approx 0.445$ and hence $s^* \approx 0.723$, in close agreement with the simulations of [6]. Note that $f_N^{\mathrm{aBS}}(2) \approx 0.209$ for all the values of N in Figure 1.

Following [1] we define the length of an s-avalanche to be ℓ if the number of consecutive steps for which the smallest fitness value stays below s is ℓ. We compute the distribution $n(\ell)$ of s-avalanche lengths for aBS and our RDP. Representative distributions are given in Figure 2. As s approaches s^* we find that $n(\ell)$ shows the power law behaviour characteristic of self-organized criticality, although there is a small but clear difference in the exponents of the two processes.

The class of RDPs defined by f_N described above is contained in a wider class of processes [3], which we now define. In this generality, an RDP is a discrete-time Markov process on the N-simplex

$$\Delta_N = \{(x_{(1)},\ldots,x_{(N)}) \colon 0 \leq x_{(1)} \leq \cdots \leq x_{(N)} \leq 1\};$$

$x_{(1)},\ldots,x_{(N)}$ are the (increasing) *order statistics* of $x_1,\ldots,x_N$. The RDP evolves according to the following Markovian rule. At each step, K of the x_k-values

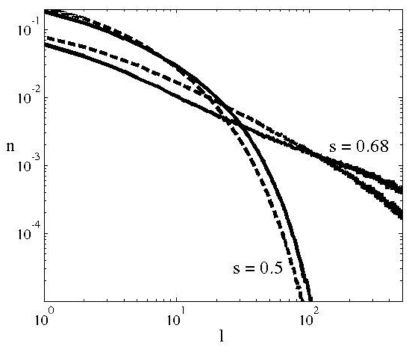

FIG. 2. Size distribution $n(\ell)$ of s avalanches in aBS (solid line) and RDP (dashed) for $s = 0.5,\ 0.68$ and $N = 1000$.

are selected, according to rank, by sampling (without replacement, according to some specified probability distribution) from $\{1, 2, \ldots, N\}$; that is, the sample from $\{1, 2, \ldots, N\}$ specifies the $x_{(k)}$ that are chosen. The chosen x_k-values are replaced by new (independent) $U[0, 1]$ values. These processes are discussed in more detail in [3].

The numerical evidence reported above leads to several interesting problems for further investigation, not least the strong suggestion that the RDP with $f_N = f_N^{\mathrm{aBS}}$ given by (1) is closely related to aBS itself. The exact relationship of the two processes remains to be characterized rigorously. If one wished to define a Markov process on Δ_N whose stationary distribution coincided with the projection onto Δ_N of the stationary distribution of aBS, a natural candidate would be a RDP with state-dependent selection distribution: instead of a single $f_N(\cdot)$ one would have a family $f_N(\,\cdot\,; x)$ of selection distributions conditioned on the state $x \in \Delta_N$. Thus, assuming it exists, one would take $f_N(\,\cdot\,; x)$ to be $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$, the stationary distribution for aBS of the right-neighbour of the smallest element *conditional* on the projection of the current state onto Δ_N being x. The fact that the numerical evidence described above suggests that one can proceed not with a state-dependent RDP based on $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$ but with the simpler RDP based on $f_N^{\mathrm{aBS}}(\,\cdot\,)$ (which is an average of the $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$) seems to point to some important underlying property of aBS itself. Two possible explanations are:

(a) $f_N^{\mathrm{aBS}}(\,\cdot\,) = f_N^{\mathrm{aBS}}(\,\cdot\,; x)$ for all x, i.e., at stationarity there is some independence between the order statistics and the permutation that maps sites to ranks; or

(b) $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$ satisfies (uniformly in x) the same asymptotic conditions as $f_N^{\mathrm{aBS}}(\,\cdot\,)$ that are central to the limit behaviour, namely (2) and (4).

The stronger fact (a) would suggest that the stationary distribution of the RDP coincides with the projection of the stationary distribution of aBS onto Δ_N, so that the two processes share the same detailed equilibrium properties. The weaker fact (b) would suffice to explain why the two processes share the same threshold and characteristic $U[s^*, 1]$ limit distribution. We remark that the distributions $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$ seem to be very difficult to evaluate numerically.

Finally, we give a very brief indication of the origin of the threshold formula (3); see [3] for details. Consider the s-counting process $N_t(s)$ defined to be the number of x_k-values in the interval $[0, s]$ after t iterations of the RDP defined by f_N. Then $N_t(s)$ is a Markov chain on the finite state-space $\{0, 1, \ldots, N\}$. The threshold s^* relates to the limiting ($t \to \infty$ then $N \to \infty$) marginal distribution of an arbitrary x_k. The probability that a randomly chosen x_k-value is less than s is $E[N_t(s)]/N$ (where E denotes expected value). Thus a natural way to define a threshold s^* is

$$s^* = \sup\{s \geq 0 : \lim_{N\to\infty} \lim_{t\to\infty} N^{-1} E[N_t(s)] = 0\};$$

the t-limit exists by Markov chain limit theory and it can be shown that the N-limit exists too, so that s^* is well defined [3].

To evaluate s^*, we compute the mean drift of $N_t(s)$:

$$\begin{aligned} E[N_{t+1}(s) - N_t(s) \mid N_t(s) = n] \\ = 2s - (1 + F_N(n))\mathbf{1}\{n > 0\}, \quad (5) \end{aligned}$$

where $F_N(n) = \sum_{k=2}^{n} f_N(k)$, $\mathbf{1}$ denotes an indicator function, and an empty sum is 0. Heuristically, for large N and large n, $F_N(n) \approx \alpha$ by (2) so that this drift is approximately $2s - 1 - \alpha$, and setting this equal to zero gives (3). One expects that the drift being zero indicates the threshold behaviour, because a positive (negative) drift would mean $N_t(s)$ increases (decreases). In this argument there are several limits involved (n, N, t all going to ∞) that need to be handled with care. We exploit techniques from Markov process theory, such as Foster–Lyapunov conditions [10], to do this: we refer to [3] for the details.

In conclusion, we have indicated how the distribution $f_N^{\mathrm{aBS}}(k)$ and the quantity α of (2) capture the build-up of correlations in Bak–Sneppen type algorithms, the threshold behaviour of which can be analysed exactly by considering the $N_t(s)$ Markov process on a countable state space.

The class of RDPs that we have introduced is of interest in its own right. Numerical evidence suggests that by choosing as parameter for the RDP an appropriate statistic ($f_N^{\mathrm{aBS}}(\,\cdot\,)$) of aBS, one can replicate the asymptotic

behaviour of aBS by the RDP, for which one can prove rigorous results more easily. The remaining analytical challenge is to clarify the relationship between aBS and the RDP. This involves at least two main parts: (i) proving the existence of the distributions f_N^{aBS} given by (1) and of the limit α defined by (2); and (ii) determining the property of aBS that allows us to use $f_N^{\mathrm{aBS}}(\,\cdot\,)$ instead of the conditional version $f_N^{\mathrm{aBS}}(\,\cdot\,; x)$. If one can make precise the connection between aBS and the RDP, one should be able to transfer rigorous results for RDPs [3] to aBS. In respect to challenge (i) above, it is interesting to note that if an explicit description of f_N^{aBS} could be obtained, one might be able to obtain an explicit formula for the threshold s^* via (2) and (3). Finally, we note that in the case of the classical BS process the same arguments apply, though now f_N is a function of two variables, $f_N(k, \ell), k \in \{2, \dots N-1\}, \ell \in \{k+1, \dots, N\}$.

MG would like to acknowledge fruitful discussions with Gregory Berkolaiko, Jack Carr, Oliver Penrose, and Michael Wilkinson.

* m.grinfeld@strath.ac.uk

† p.a.knight@strath.ac.uk

‡ andrew.wade@strath.ac.uk

[1] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).

[2] H. J. Jensen, *Self-Organized Criticality* (Cambridge University Press, 1998).

[3] M. Grinfeld, P. A. Knight, and A. R. Wade, (in preparation).

[4] This paper has 868 citations according to Google Scholar accessed on 8/11/10.

[5] R. Meester and D. Znamenski, Ann. Probab. **31**, 1986 (2003).

[6] G. J. M. Garcia and R. Dickman, Physica A **342**, 516 (2004).

[7] H. Flyvbjerg, K. Sneppen, and P. Bak, Phys. Rev. Lett. **71**, 4087 (1993).

[8] J. de Boer, B. Derrida, H. Flyvbjerg, A. D. Jackson, and T. Wettig, Phys. Rev. Lett. **73**, 906 (1994).

[9] G. L. Labzowsky and Y. M. Pis'mak, Phys. Lett. A **246**, 377 (1998).

[10] G. Fayolle, V. A. Malyshev, and M. V. Menshikov, *Topics in the Constructive Theory of Countable Markov Chains* (Cambridge University Press, 1995).